Amendments to the Claims:

This listing of the claims will replace all prior versions and listings of the claims in the application:

Listing of Claims:

1. (Currently Amended) A noise reduction system ~~with~~ including an audio-visual user interface therein ~~, said system being specially adapted for running an application~~ for combining visual features ~~($o_{v,nT}$)~~ extracted from a digital video sequence ~~($v(nT)$) showing the face of a speaker (S_i)~~ with audio features ~~($o_{a,nT}$)~~ extracted from an analog audio sequence ~~($s(t)$), wherein said audio sequence ($s(t)$) can include~~ including background noise in ~~the~~ an environment of ~~said~~ a speaker ~~(S_i)~~, said noise reduction system ~~(200b/c)~~ comprising:

[[-]] audio sequence detection means ~~(101a, 106b)~~ for detecting said analog audio sequence; ~~and~~

audio feature extraction and analysis means for analyzing said analog audio sequence ~~($s(t)$),~~ and extracting said audio features therefrom;

[[-]] video sequence detection means ~~(101b')~~ for detecting said video sequence ~~($v(nT)$), and~~ ;

[[-]] visual feature extraction and analysis means ~~(104a+b, 104'+104'')~~ for analyzing the detected video sequence ~~signal ($v(nT)$),~~ and extracting said visual features therefrom;

~~wherein~~ a noise reduction circuit ~~(106) of said noise reduction system is adapted~~ configured to separate [[the]] a speaker's voice from said background noise ~~($n'(t)$)~~ based on a combination of derived speech characteristics ~~($o_{av,nT} := [o_{a,nT}^T, o_{v,nT}^T]^T$)~~ and ~~outputting~~ configured to output a speech activity indication signal ~~($\hat{s}_i(nT)$) which is obtained by~~ comprising a combination of speech activity estimates supplied by said audio feature extraction and analysis means and said visual feature extraction and analysis ~~analyzing~~ means ~~(106b, 104a+b, 104'+104''),~~ ; and ~~**characterized by**~~

a multi-channel acoustic echo cancellation unit ~~(108) being specially adapted~~ configured to perform a near-end speaker detection and double-talk detection algorithm based on ~~acoustic-phonetic~~ the speech characteristics derived by said audio feature extraction and

~~analyzing~~ analysis means ~~(106b)~~ and said visual feature extraction and ~~analyzing~~ analysis means ~~(104a+b, 104'+104'')~~.

2. (Currently Amended) A noise reduction system according to claim 1, ~~characterized by~~ further comprising:

means ~~(SW)~~ for switching off an audio channel ~~in case the actual level of~~ if said speech activity indication signal ~~($\hat{s}_i(nT)$)~~ falls below a predefined threshold value.

3. (Currently Amended) A noise reduction system according to ~~anyone of the claims 1 or 2, characterized in that~~ claim 1, wherein said audio feature extraction and ~~analyzing~~ analysis means ~~(106b) is~~ comprises an amplitude detector.

4. (Currently Amended) A near-end speaker detection method for reducing ~~the~~ noise ~~level of~~ in a detected analog audio sequence ~~($s(t)$)~~, said method ~~being characterized by the following steps~~ comprising:

[[-]] ~~subjecting (S1)~~ converting said analog audio sequence ~~($s(t)$) to an analog to~~ into a digital ~~conversion,~~ audio sequence;

[[-]] calculating ~~(S2) the~~ a corresponding discrete signal spectrum ~~($S(k\cdot\Delta f)$)~~ of the ~~analog-to-~~digital-~~converted~~ audio sequence ~~($s(nT)$)~~ by performing a Fast Fourier Transform (FFT)[[,]] ;

[[-]] detecting ~~(S3) the~~ a voice of ~~said~~ a speaker ~~(S_i)~~ from said discrete signal spectrum ~~($S(k\cdot\Delta f)$)~~ by analyzing visual features ~~($o_{v;nT}$)~~ extracted from a ~~simultaneously with the recording of the analog audio sequence ($s(t)$) recorded~~ video sequence ~~($v(nT)$) tracking the~~ associated with the audio sequence and including current ~~location~~ locations of ~~the speaker's~~ face, lip movements and/or facial expressions of the speaker ~~(S_i)~~ in ~~subsequent~~ a sequence of images in the video sequence [[,]] ;

[[-]] estimating ~~(S4) the~~ a noise power density spectrum ~~($\Phi_{nn}(f)$)~~ of [[the]] statistically distributed background noise ~~($\tilde{n}(t)$)~~ based on [[the]] ~~result of the speaker detection step (S3),~~ detection of the voice of the speaker;

[[-]] subtracting ~~(S5)~~ a discretized version ~~($\widetilde{\Phi}_{nn}(k \cdot \Delta f)$)~~ of the estimated noise power density spectrum ~~($\widetilde{\Phi}_{nn}(f)$)~~ from the discrete signal spectrum ~~($S(k \cdot \Delta f)$)~~ of the ~~analog-to-~~digital-~~converted~~ audio sequence ~~($s(nT)$),~~ to obtain a difference signal; and

[[-]] calculating ~~(S6) the~~ a corresponding discrete time-domain signal ~~($\hat{s}_i(nT)$)~~ of the obtained difference signal by performing an Inverse Fast Fourier Transform (IFFT)~~, thereby yielding a discrete version of the~~ to provide a recognized speech signal.

5. (Currently Amended) A near-end speaker detection method according to claim 4, ~~characterized by the step of~~ further comprising:

~~conducting (S7)~~ performing a multi-channel acoustic echo cancellation algorithm which models echo path impulse responses by means of adaptive finite impulse response (FIR) filters and subtracts echo signals from the analog audio sequence ~~($s(t)$)~~ based on acoustic-phonetic speech characteristics derived by an algorithm for extracting the visual features ~~($o_{v,nT}$)~~ from [[a]] the video sequence ~~($v(nT)$) tracking the location~~ associated with the audio sequence and including the locations of [[a]] the ~~speaker's~~ face, lip movements and/or facial expressions of the speaker ~~(S_i)~~ in ~~subsequent~~ a sequence of images in the video sequence.

6. (Currently Amended) A near-end speaker detection method according to claim 5, ~~characterized in that~~ wherein said multi-channel acoustic echo cancellation algorithm performs a double-talk detection procedure.

7. (Currently Amended) A near-end speaker detection method according to ~~anyone of the claims 4 to 6~~ claim 4, ~~characterized in that~~ wherein said acoustic-phonetic speech characteristics are based on [[the]] detecting opening of a ~~speaker's~~ mouth of the speaker as an estimate of [[the]] acoustic energy of articulated vowels [[or]] and/or diphthongs, ~~respectively,~~ detecting rapid movement of the ~~speaker's~~ lips of the speaker as a hint to labial or labio-dental consonants, ~~respectively, and~~ and/or detecting other ~~statistically~~

~~detected~~ phonetic characteristics ~~of an association between~~ associated with position and movement of the lips ~~and the~~ and/or voice ~~and~~ and/or pronunciation of said speaker ~~(S_i)~~.

8. (Currently Amended) A near-end speaker detection method according to ~~anyone of the claims 4 to 7, **characterized by**~~ claim 4, wherein detecting the voice of said speaker comprises:

~~a learning procedure used for enhancing the step of~~ detecting ~~(S3)~~ the voice of said speaker ~~(S_i)~~ from the discrete signal spectrum ~~($S(k \cdot \Delta f)$)~~ of the ~~analog-to-~~digital-~~converted version ($s(nT)$) of an analog~~ audio sequence ~~($s(t)$)~~ using a learning procedure by analyzing the visual features ~~($o_{v,nT}$)~~ extracted from ~~a simultaneously with the recording of the analog audio sequence ($s(t)$) recorded~~ the video sequence ~~($v(nT)$) tracking the~~ associated with the audio sequence and including the current ~~location~~ locations of the ~~speaker's~~ face, lip movements and/or facial expressions of the speaker ~~(S_i)~~ in ~~subsequent~~ a sequence of images in the video sequence.

9. (Currently Amended) A near-end speaker detection method according to ~~anyone of the claims 4 to 8, **characterized by** the step of~~ claim 4, further comprising:

correlating ~~(S8a)~~ the discrete signal spectrum ~~($S_\tau(k \cdot \Delta f)$)~~ of a delayed version ~~($s(nT-\tau)$)~~ of the ~~analog-to-~~digital-~~converted~~ audio signal ~~($s(nT)$)~~ with an audio speech activity estimate obtained by [[an]] amplitude detection ~~(S8b)~~ of [[the]] a band-pass-filtered discrete signal spectrum ~~($S(k \cdot \Delta f)$), thereby yielding~~ to provide an estimate ~~($\widetilde{S}_i(f)$)~~ for [[the]] a frequency spectrum ~~($S_i(f)$)~~ corresponding to [[the]] a signal ~~($s_i(t)$)~~ which represents ~~said speaker's~~ a voice of said speaker as well as an estimate ~~($\widetilde{\Phi}_{nn}(f)$)~~ for the noise power density spectrum ~~($\Phi_{nn}(f)$)~~ of the statistically distributed background noise ~~($n'(t)$)~~.

10. (Currently Amended) A near-end speaker detection method according to claim 9, ~~**characterized by** the step of~~ further comprising:

correlating ~~(S9)~~ the discrete signal spectrum ~~($S_\tau(k \cdot \Delta f)$)~~ of [[a]] the delayed version ~~($s(nT-\tau)$)~~ of the ~~analog-to-~~digital-~~converted~~ audio signal ~~($s(nT)$)~~ with a visual speech activity

estimate taken from a visual feature vector ~~($o_{v,t}$)~~ supplied by the visual feature extraction and analyzing means ~~(104a+b, 104'+104''), thereby yielding~~ to provide a further estimate ~~($\tilde{S}_i'(f)$)~~ for updating the estimate ~~($\tilde{S}_i(f)$)~~ for the frequency spectrum ~~($S_i(f)$)~~ corresponding to the signal ~~($s_i(t)$)~~ which represents said speaker's voice as well as a further estimate ~~($\tilde{\Phi}_{nn}'(f)$)~~ for updating the estimate ~~($\tilde{\Phi}_{nn}(f)$)~~ for the noise power density spectrum ~~($\Phi_{nn}(f)$)~~ of the statistically distributed background noise ~~($n'(t)$)~~.

11. (Currently Amended) A near-end speaker detection method according ~~anyone of the claims 9 or 10, **characterized by** the step of~~ to claim 9, further comprising:

adjusting ~~(S10) the~~ cut-off frequencies of a band-pass filter ~~(204)~~ used for filtering the discrete signal spectrum ~~($S(k\cdot\Delta f)$)~~ of the ~~analog-to-~~digital-~~converted~~ audio ~~signal ($s(t)$) dependent~~ sequence based on [[the]] a bandwidth of the estimated ~~speech signal~~ frequency spectrum ~~($\tilde{S}_i(f)$)~~.

12. (Currently Amended) A near-end speaker detection method according to ~~anyone of the claims 4 to 8, **characterized by** the steps of~~ claim 4, further comprising:

[[-]] adding ~~(S11a)~~ an audio speech activity estimate obtained by [[an]] amplitude detection of [[the]] a band-pass-filtered discrete signal spectrum ~~($S(k\cdot\Delta f)$)~~ of the ~~analog-to-~~digital-~~converted~~ audio ~~signal ($s(t)$)~~ sequence to a visual speech activity estimate taken from a visual feature vector [[($o_{v,t}$)]] supplied by said visual feature extraction and analyzing means ~~(104a+b, 104'+104''), thereby yielding~~ to provide an audio-visual speech activity estimate,

[[-]] correlating ~~(S11b)~~ the discrete signal spectrum ~~($S(k\cdot\Delta f)$)~~ with the audio-visual speech activity estimate~~, thereby yielding~~ to provide an estimate ~~($\tilde{S}_i(f)$)~~ for [[the]] a frequency spectrum ~~($S_i(f)$)~~ corresponding to [[the]] a signal ~~($s_i(t)$)~~ which represents ~~said speaker's~~ a voice of said speaker as well as an estimate ~~($\tilde{\Phi}_{nn}(f)$)~~ for the noise power density spectrum ~~($\Phi_{nn}(f)$)~~ of the statistically distributed background noise ~~($n'(t)$)~~ ; and

[[-]] adjusting ~~(S11c) the~~ cut-off frequencies of a band-pass filter ~~(204)~~ used for filtering the discrete signal spectrum ~~($S(k\cdot\Delta f)$)~~ of the ~~analog-to-~~digital-~~converted~~ audio ~~signal~~

~~($s(t)$) dependent~~ sequence based on [[the]] a bandwidth of the estimated ~~speech signal~~ frequency spectrum ~~($\widetilde{S}_i(f)$)~~.

13. (Currently Amended) A telecommunication system, comprising:

~~Use of a noise reduction system (200b/c) according to anyone of the claims 1 to 3 and a near-end speaker detection method according to anyone of the claims 5 to 13 for~~

a video-enabled phone;

a video-telephony based application ~~in a telecommunication system~~ running on [[a]] the video-enabled phone ~~with~~ ; and

a ~~built-in~~ video camera ~~(101b')~~ built-in to the video-enabled phone and pointing at [[the]] a face of a speaker ~~(S_i)~~ participating in a video telephony session,

wherein said video-telephony based application comprises:

audio sequence detection means for detecting an analog audio sequence;

audio feature extraction and analysis means for analyzing said analog audio sequence and extracting said audio features therefrom;

video sequence detection means for detecting said video sequence;

visual feature extraction and analysis means for analyzing the detected video sequence and extracting said visual features therefrom;

noise reduction means for separating a speaker's voice from said background noise based on a combination of derived speech characteristics and outputting a speech activity indication signal comprising a combination of speech activity estimates supplied by said audio feature extraction and analysis means and said visual feature extraction and analysis means; and

multi-channel acoustic echo cancellation means for performing a near-end speaker detection and double-talk detection algorithm based on the speech characteristics derived by said audio feature extraction and analysis means and said visual feature extraction and analysis means.

14. (Currently Amended) A telecommunication device equipped with an audio-visual user interface~~, characterized by~~ and including the noise reduction system ~~(200b/c)~~ according to ~~anyone of the claims 1 to 3~~ claim 1.

15. (New) A telecommunication system configured to perform the near-end speaker detection method of claim 4.